iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4126

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

An Organized Anomaly Recognition Scheme for WSNs

Dr. Syeda Gauhar Fatima[1], Syeda Kausar Fatima[2], Syed Mohd Ali[3]
[1]*Prof. DCET, Hyderabad, Associate Prof. SCET, Hyderabad, Research Scholar , JNTUH*

Abstract: ***Wireless sensor networks are significant networks for gathering environmental statistics and observing phenomena. Irregularities initiated by hardware and software faults, strange actions, and malicious attackers disturb the reliability of information collected through such networks. Hence, abnormality recognition procedure is an essential phase in constructing sensor networks to guarantee data excellence for correct assessment. This paper presents a novel scattered working anomaly recognition scheme that processes the variation of sensor estimations in major module space. This model allocates the discovery process above the network to reduce energy consumption whereas confirming high recognition value. The recognition effectiveness and proficiency of this scheme are shown through trials on real world dataset from Sensor scope system project. Experimental outcomes reveal this scheme accomplished high recognition level with comparatively little incorrect positive rates compared to an existing scheme.***

Keywords: Wireless Sensor Networks; Anomaly Recognition; Distributed Processing; Principal Component Analysis; Principal Component Classifier.

I. INTRODUCTION

Wireless sensor networks are small sized, low power, energy limited, and multifunctional devices named sensors that are used to gather data from an environment or monitor an occurrence [1]. The limited resources such as processing capabilities and energy increase the possibility of these networks to be susceptible to variety of misbehaviors or anomalies. Data anomaly is defined in [2] as, “an observation that appears to be inconsistent with the reminder of a dataset”. Hardware and software faults, observations errors, unusual events, or intrusions are possible causes for sensor data anomalies. Therefore, to assure the integrity of sensor observations and identify important interested events in the monitored phenomena, detection of anomalies should be considered for the design of any sensor network system. Distribution of anomaly detection process over nodes in WSN has been proposed as a solution to reduce the computational load of detection methods in nodes. However, another factor should be considered when designing any distributed detection model which is the communication overhead caused by transmitting large amounts of data between nodes. Some distributed anomaly detection models have been designed to consider the communication overhead such as [3-6]. The concept of sharing in these works depend on on that each node implements the abnormality detection technique by its own resources and then transmits only a summary of local detection model to the significant location such as cluster head or base station in order to construct a global detection model. The global model is then returned to the nodes for using it in anomaly detection of subsequent observations. A distributed deviation detection model in WSN was described in [4] to avoid the unnecessary communication and computational effort. This model was designed based on a non-parametric statistical technique called kernel density estimator. The hierarchical structure which is always used to implement any distributed solution was not adopted for this model. Instead, the hierarchical structure was emulated by assigning each group of low capacity sensors to one of some limited more powerful sensors based on spatial proximity. An In-network outlier detection algorithm was proposed in [5] and aimed at reducing the communication overhead results from massive data transmission in WSN. This algorithm was based on distance similarity to find the global anomalies in WSN. In this model, each node applies the distance similarity measure to find the anomalies and broadcast its result to its neighbours. Other nodes perform the same task until all nodes agree on a common decision on anomalies. Broadcast communication lead to high energy consumption. The problem is worsen when deal with large scale WSN.
A Quarter Sphere Support Vector Machine (QSSVM) based distributed anomaly detection model was proposed in [6] to detect the anomalous observations in sensors. The technique of applying this model is as follows:

1) Each sensor node runs the QSSVM on its own data to find the local anomalous observations and the parameter R_j of the QSSVM that distinct normal observations from anomalous.
2) Every single sensor directs the local parameter R_j to the parent node in the configuration.
3) The parent node gathers the radii of the children nodes and constructs the global radius R_m.
4) The parent returns the global radius parameter R_m to its children nodes to detect anomalies in their data by means of this radius.

©IJRASET: All Rights are Reserved

Range of distributed prototypes was considered based on the QSSVM [3, 7-10]; nevertheless, the common inadequacy of these replicas is the forbidden computational cost of applying the QSSVM on the inadequate sensor resources. Furthermore, the need of SVM on some constraints significantly disturbs the detection efficiency. These factors must be chosen for each WSN application and hence upset the pragmatism of the solution. The authors of [15] proposed statistical based outlier detection methods based on temporal and spatial correlations of sensor data. The performance of these approaches is extremely dependent on certain statistical parameters that disturb the usability of these methods for operational detection. In this paper, a new distributed accessible anomaly recognition scheme is considered by means of the one class principal component classifier (OCPCC) previously proposed in [11] to detect anomalies as they happened. The modification between the scheme proposed in this paper and the model in [11] is the selection of detection threshold and the variant of PCA algorithm used. In [11], the threshold was chosen by clustering the dissimilarity vector of training data via automated clustering threshold method which has high computational cost to be used for online detection. In addition, the called candid-covariance free incremental principal component analysis (CCIPCA) algorithm which was originally proposed in [12] and utilized for data reduction for WSN in [13] was used as a core for the plan of the proposed scheme in this paper. Data samples chosen from sensor scope GSB [14] dataset were used to authenticate the detection usefulness of this scheme in terms of detection rate and false alarm rate. An evaluation with an existing distributed model in [15] was directed to show the advantages of the proposed scheme.

The rest of paper is organized as follows: section 2 details out the plan of the proposed model. Section 3 provides the experimental results and discussion. Section 4 concludes the paper and recommends some future research guidelines.

II. PROPOSED SCHEME

Two main stages are involved in the design of the proposed scheme which are training stage and detection stages. Details of each stage are given in the following sections.

A. *Training Stage*

In this stage, the normal observations are gathered at every sensor to find out the local normal model (LNM) and send it to the cluster head (CH) for constructing the global normal model (GNM). The procedure of building the LNM is as follows: the training normal observations S_{Train} are standardized by the mean value (μ) and the standard deviation value (σ) of training observations. The eigenvector (V_i), and eigenvalues (λ_i) matrices are then calculated from the standardized observations. The projection of training observations on the principal component space Y_i is then computed by Eq. (1):

$$Y_i = {}^{S}{}_{Train}(i) \times V(i) \qquad (1)$$

The number of principal component which depends on the application is determined and the D_{train} dissimilarity measure is computed for training observations using Eq. (2).

$$D_{train} = \sum \frac{i}{\lambda} Y \qquad (2)$$

In this paper, the detection threshold that represents the local normal model (LNM) is chosen to be the maximum *MaxDiss* and minimum *MinDiss* bounds of the dissimilarity vector D_{train}. The LNM parameters are then sent to CH to construct the GNM. Different strategies are used to build the GNM at CH; however, the details of such strategies are not the subject of this paper.

B. *Detection stage*

In this stage, each node tests every observation using the GNM built in training stage. Some other parameters which include the mean (μ) and standard deviation (σ) parameters, eigenvector (V_i) and eigenvalues (λ_i) matrices, are also used in this stage. The GNM model is composed of (*MaxDiss, MinDiss)* values calculated globally at CH. Each observation is identified as either normal or anomalous by comparing its projection on the space with the detection thresholds specified in the GNM model. The score of projecting every new observation on the principal component space is computed using the parameters of the normal profile that were obtained from training stage as in Eq. (3): $Y_i {}^{S}{}_{Test} V$

The measure of dissimilarity (D_{test}) for every new observation is computed using Eq. (4).

$$= \sum \frac{y}{\lambda} \quad 2D_{test}$$

$= \quad \times$

©IJRASET: All Rights are Reserved

At the end, the value of D_{test} parameter is compared against the GNM values to classify each observation as normal or anomalous using the following criterion:

$$(Dtest > MaxDiss) Or (Dtest < MinDiss) \rightarrow Anomalous$$

$$else \rightarrow Normal$$

III. EXPERIMENTAL RESULTS AND EVALUATION

In this section, the experimental results of the proposed scheme is based on real world data set obtainable and estimated by comparing it to a recent existing anomaly detection methd from the literature.

A. Datasets

Grand St. Bernard (GSB) [14] is one of sensor scope project deployment dataset was gathered using WSN deployment.The network is formed of 23 sensors that record metrological environmental data that include; (ambient and surface) temperature, and humidity. 2 clusters were formed in this deployment in which the small cluster has 5 sensors and the big cluster has 18 sensors. The observations of the small cluster which have 5 sensors namely, Node25, Node-28, Node-29, Node-31 and Node-32 were used to evaluate the proposed model. In our experiments, the ambient temperature observations at the time period 6am of data recorded were used. The observations are labeled using the Mahalanobis distance measure as used in [15].

B. Results and Discussion

We have tested our proposed model on dataset samples extracted from nodes 25,28,29,31 and 32. The hierarchical WSN structure was adopted in which one of the above nodes act as CH that receives the LNM from other nodes in the cluster and constructs the GNM that is used by other cluster nodes for detection. The detection rate (DR) and false alarm rates (FPR) which represent detection effectiveness are reported in Table 1.

Table 1. Detection rate and false alarm rate of the proposed model

	Nodes					
	N25	N28	N29	N31	N32	Average
DR (%)	100	91	100	89	100	96
FPR (%)	16	0	2	0	0	7.2

Table 1 reports the results of the proposed scheme for each node. The results show that an average of 96% detection rate was achieved by the scheme while 7.2% false positives were produced from misclassification of some normal observations as anomalies. A comparison with the statistical temporal based outlier detection (TOD) model proposed in [15] using the same data samples and same data labeling approach is shown in Table 2.

Table 2. A comparison with the TOD model proposed in [15]

	SW=15	SW=30	SW=48	SW=60	Our model
DR (%)	82.86	82.86	82.86	82.86	96
FPR (%)	13.3	13.48	11.67	10.13	7.2

Table 2 reports the results of the TOD model with different smoothing windows values. It is clearly shown that our proposed model outperforms the TOD model in terms of detection rate and false positive rate. To evaluate the efficiency of the model, the computational complexity and communication overhead are considered. The computational complexity incurred by our model is *O(N)* where *N* is the number of observed variable in the phenomenon. This complexity is the computational complexity of CCIPCA algorithm. The additional calculations involved in the model are fixed. Since the LNM sent by each node has a fixed size, the communication overhead is a function of the number of nodes *K* in each cluster; therefore, the communication overhead is *O(K)*.

©IJRASET: All Rights are Reserved

For the TOD model, there is no communication overhead as the method is implemented locally but the computation cost is $O(c.\ d\ .p)$ where d is the number of variables, c is the number of original observations to be modeled, and p is the fitting parameter of the AR model used by the TOD.

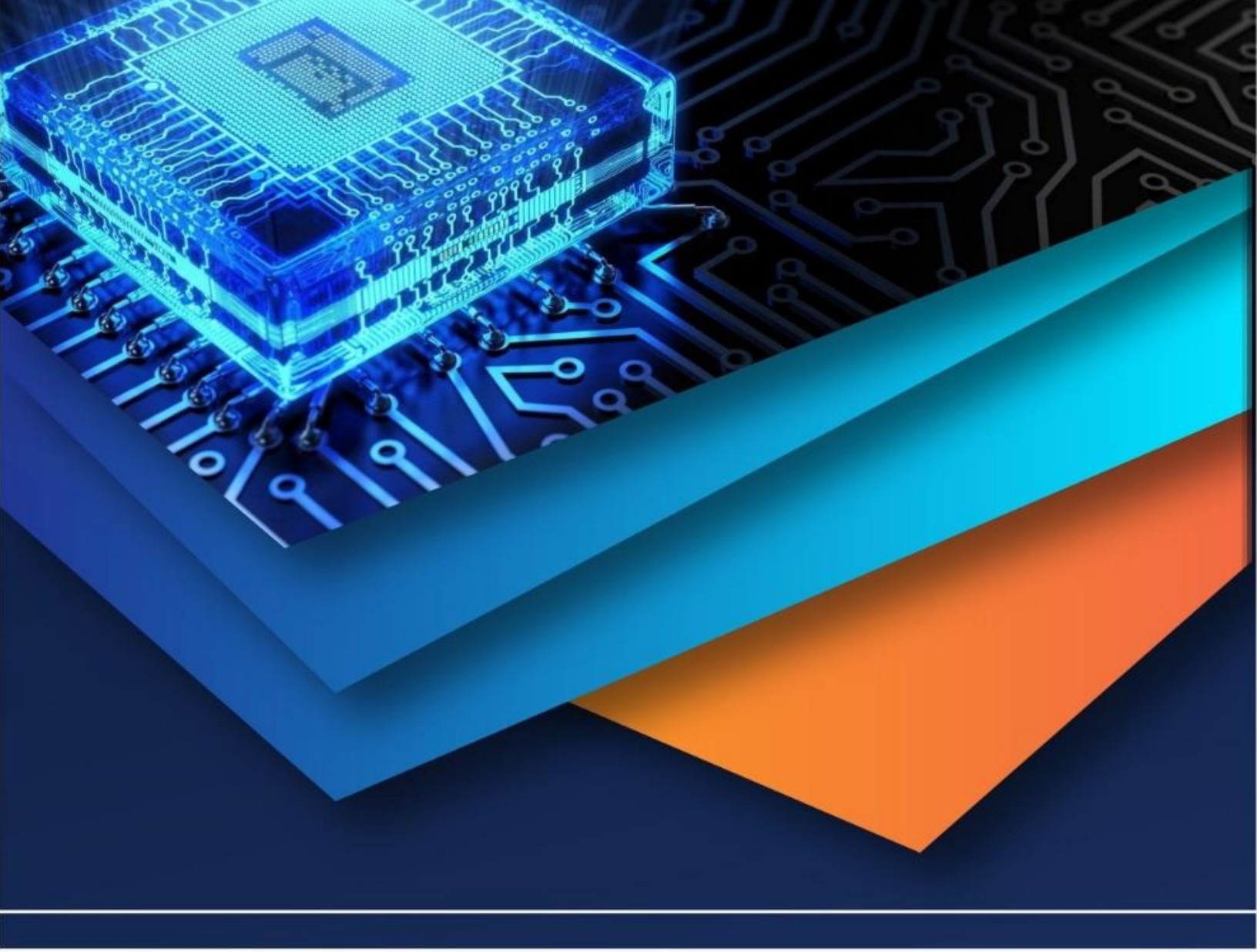
IV. CONCLUSION

This paper introduces the preliminaries and initial experimental results of new distributed anomaly detection scheme for WSNs based on the calculation of the dissimilarity between sensor observations in the principal component space. Experimental results and comparison with recent existing work indicate that the scheme is promising in terms of achieving high detection effectiveness while efficiently utilizing the limited network resources. Further experiments are needed to investigate the performance of the scheme with multivariate data and additional real world datasets.

REFERENCES

[1] Akyildiz I.F, Su W, Sankarasubramaniam Y , Cayirci E. Wireless sensor networks: a survey. Computer Networks 2002. 38(4); 393-422.
[2] Hodge V, Austin J. A Survey of Outlier Detection Methodologies. Artif. Intell. Rev 2004. 22(2); 85-126.
[3] Zhang Y, Meratnia N, Havinga P. An online outlier detection technique for wireless sensor networks using unsupervised quarter-sphere support vector machine. In International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 2008.
[4] Palpanas T, Papadopoulos D, Kalogeraki V, Gunopulos D. Distributed deviation detection in sensor networks. SIGMOD Rec., 2003. 32(4);77-82.
[5] Branch J, Szymanski B, Giannella C, Wolff R,Kargupta H. In-Network Outlier Detection in Wireless Sensor Networks. In 26th IEEE International Conference on Distributed Computing Systems, 2006.
[6] Rajasegarar S, Leckie C, Palaniswami M, Bezdek J. Quarter Sphere Based Distributed Anomaly Detection in Wireless Sensor Networks. in IEEE International Conference on Communications, ICC '07. 2007.
[7] Shahid N, Naqvi I.H, Qaisar S.B. Quarter-Sphere SVM: Attribute and Spatio-Temporal correlations based Outlier & Event Detection in wireless sensor networks. In 2012 IEEE Wireless Communications and Networking Conference (WCNC),2012.
[8] Takianngam S, Usaha W. Discrete Wavelet Transform and One-Class Support Vector Machines for anomaly detection in wireless sensor networks. In 2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), 2011.
[9] Rajasegarar S, Leckie C, Bezdek, J.C, Palaniswami M. Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks. IEEE Transactions on Information Forensics and Security, 2010. 5(3); 518-533.
[10] Zhang Y, Meratnia N, Havinga P.J. Ensuring high sensor data quality through use of online outlier detection techniques. International Journal of Sensor Networks, 2010. 7(3); 141-151.
[11] Rassam M.A, Zainal A, Maarof M.A. One-Class Principal Component Classifier for Anomaly Detection in Wireless Sensor Network. In 2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN), 2012; 271-276.
[12] Juyang W, Yilu Z, Wey-Shiuan, H. Candid covariance-free incremental principal component analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003. 25(8); 1034-1040.
[13] Rassam M.A, Zainal A, and Maarof M.A. An Adaptive and Efficient Dimension Reduction Model for Multivariate Wireless Sensor Networks Applications. Applied Soft Computing, 2013. 13 (4): 1878-1996.
[14] GSB, Grand-St-Bernard (GSB) dataset. http://lcav.epfl.ch/cms/lang/en/pid/86035. 2007.
[15] Zhang Y, Hamm N, Meratnia N, Havinga P. Statistics-based outlier detection for wireless sensor networks. International Journal of Geographical Information Science, 2012. 26(8);1373-1392.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)